DOCUMENT RESUME

ED 072 833 LI 004 210

AUTHOR Tell, Bjorn; Gluchowicz, Zofia
TITLE Progress Report of Computerized I&D Services at the Royal Institute of Technology, Stockholm.
INSTITUTION Royal Inst. of Technology, Stockholm (Sweden).
REPORT NO TRITA-LIB-4014
PUB DATE Oct 72
NOTE 30p.; (4 References)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Data Bases; *Electronic Data Processing; Foreign Countries; *Information Retrieval; *Information Services; *Information Systems; On Line Systems; Search Strategies
IDENTIFIERS ERIC Data Base; Scientific and Technical Information; Selective Dissemination of Information; *Sweden

ABSTRACT

During the five years of activities the documentation centre at the Royal Institute of Technology has established itself as an information centre in the fields of science and technology. The SDI service is now well implemented and its activities are used and appreciated by scientists, research workers and engineers at the universities, research institutions and in the industrial communities. Techniques for on-line SDI-query formulation and query alternation adaptive to user feedback are under development. The on-line connection to the NASA:s Recon system in Darmstadt enables us to make retrospective searches in interactive mode. Research is going on for linking up the Swedish network for Library Information system - LIBRIS - with international data banks with the objective to achieve a comprehensive information retrieval system for the whole country. (Author)

RAPPORT-TRITA-LIB-4014

PROGRESS REPORT OF COMPUTERIZED I&D SERVICES AT THE ROYAL INSTITUTE OF TECHNOLOGY, STOCKHOLM

Björn Tell and Zofia Gluchowicz

Kungl. Tekniska Högskolan
S-100 44 STOCKHOLM

OKTOBER 1972

TRITA LIB-4014
Progress report of computerized I&D services at the Royal Institute of Technology, Stockholm

1. INTRODUCTION

The Swedish government has taken an active interest in developing a policy for economic growth. In 1967 it launched a program for the promotion of technological development and industrial growth, in which a plan for the development of scientific and technical information was included. The government was especially interested in studying the viability of mechanized information services in the field of science and technology, and the utility they could offer to users in research and industry. The Royal Institute of Technology library was chosen as the responsible agent for the establishment of a mechanized service for users in science, industry and education.

The requirements for the computer operation of a service had been thoroughly studied during Tell's years as department manager of the Swedish nuclear establishment, AB Atomenergi. Then, in 1967 the Institute library received the first grant of Sw.Cr. 80,000($16,000) to initiate a computerized service in the field of mechanical engineering. During the years the scope has extended and the grant has increased, and it has now stabilized around 1 Million Sw.Cr. outside the ordinary budget of the library. Half of that sum goes to the salaries for documentalists who have been added to the library staff. Thus, the fundamental requirements for staff and funds have been fulfilled by the new policy.

2. THE BASIC TASKS OF AN INFORMATION RETRIEVAL SERVICE

A computerized information system has to perform a number of basic functions, such as

- Entering various types of data
- Formatting, abbreviating and coding of data
- Processing information, i.e. searching, matching, sorting etc.
- Producing standardized or specialized types of output, e.g. bibliographies, indexes, SDI etc.
- Answering specific, one-time requests, i.e. retrospective searches
- Reacting to various errors
- Relating to other information systems

3. THE ORGANIZATION OF A NEW COMPUTERIZED SERVICE

In order to start a computerized service the best choice, at least at that time, seemed to be a current awareness service- SDI - Selective Dissemination of Information. SDI is a system developed by late Hans Peter Luhn at IBM in 1959 for alerting participants about new publications such as journal articles, reports, conference papers etc. The acronym SDI has the special connotation that the process makes use of a computer. This is possible when the references to the literature are stored on machine-readable media.

The system should be so designed that the selection and announcement of current documents should have a high probability of interest to the individual user. For this purpose the user must submit and routinely modify his "interest profile" which serves as basis for the computer matching of stored profiles against titles of indexing terms in the references.

In order to keep the interest alive on the part of the participants, the SDI service must be prepared to offer a comprehensive coverage of the literature, and a backup of pertinent material. One of the major tasks in the expansion of the library service during the past five years has been to answer the incoming queries, resulting in profiles, as broadly as possible, and install new bibliographic data bases in case they could contribute to the broadening of the subject coverage.

By using a general information retrieval system (Tell 1), it has been possible to include additional files in the service, so that the search procedure and output routines can be the same. By a "general" system we mean that it can make use of all the keys, tools and techniques for selecting references in response to a search request, e.g. classification schemes, keywords, words in titles or abstracts, author or author affiliation names, citations etc., all of which can be used in traditional, manual searches.

SDI-system at the Royal Institute of Technology, Stockholm. Databases. 1972.

1. ISI — Science Citation Index Source Data Tape from the Institute for Scientific Information (USA), containing interdiciplinary information from the most frequently cited journals in science and technology, stores about 400 000 references a year.

2. MechEn — Mechanical engineering from the Royal Institute of Technology, Stockholm covers the literature in mechanical engineering and metallurgy and stores about 40 000 references a year.

3. CAC — Chemical Abstracts Condensates from Chemical Abstracts Service (USA) stores about 340 000 references a year to literature in the field of chemistry.

4. Inspec — Information Service in Physics, Electrotechnology and Computers & Control from the Institution of Electrical Engineers (U.K.) in collaboration with the Institute of Electrical and Electronics Engineers (USA). This is the most comprehensive information system within the fields given in the title and it stores about 120 000 references a year.

5. Metadex — Metals Abstracts Index Tapes from the American Society for Metals in collaboration with the Institute of Metals (U.K.) stores about 24 000 references a year to literature in the field of metallurgy.

6. GRA — Government Reports Announcements from the National Technical Information Service (NTIS), USA. This information system stores about 40 000 references a year to reports on USA federal sponsored research in the fields of science and technology.

7. COMPENDEX — Computerized Engineering Index from Engineering Index Inc. (USA) covers the literature in engineering and technology and stores about 72 000 references a year.

8. NSA — Nuclear Science Abstracts from the United States Atomic Energy Commission stores about 50 000 references a year. Literature searching on the NSA database is carried out in close collaboration with AB Atomenergi.

9. ABIPC — Abstract Bulletin of the Institute of Paper Chemistry from the Institute of Paper Chemistry (USA) stores about 10 000 references a year to recently published articles, patents, and theses in the field of pulp and paper chemistry and technology.

10. WOOD — WOOD from the Swedish Forest Products Research Laboratory and the Royal Institute of Technology Library, Stockholm stores about 15 000 references a year in the field of wood technology.

11. FSTA — Food Science and Technology Abstracts from the International Food Information Service (Germany) covers the literature in food science and chemistry and stores about 12 000 references a year.

12. ERIC — ERIC Master Files from the Educational Resources Information Center (USA) stores about 30 000 references a year to reports and articles, and other publications in pedagogics and modern educational science.

13. NYFLI — Accession List from the Royal Institute of Technology Library, Stockholm annually stores about 7000 titles to literature aquired by the libraries of AB Atomenergi, Chalmer's Institute of Technology,and the Royal Institute of Technology.

14. STAR — Scientific and Technical Aerospace Reports from National Aeronautics and Space Administration (USA) stores about 45 000 references a year to reports from all fields connected with aeronautics and space technology.

15. IAA — International Aerospace Abstracts from the American Institute of Aeronautics and Astronautics (USA) stores about 50 000 references a year to journals, meetings, patents, and other literature in the same field as STAR.

Databases 14 and 15 are searched at the ESRO documentation centre.

5. IMPLEMENTATION OF THE DATA BASES INTO ABACUS & VIRA

The basic approach employed has been to use a general processing format into which a record of a particular output of different files can be converted by a reformatting program so that its records can be searched. The success of this pragmatic approach to the compatibility problem of various tape formats greatly depends upon the hospitality of the search record format. The ABACUS was designed in 1966, before the MARC pilot program and the interchange format reflected in International Standard ISO/DIS 2709 which is foreseen as the standard for UNISIST. However, the ABACUS record has many characteristics in common with MARC and ISO. A directory to the whole record maps out the record length, the data elements present, and the number of characters in each element. The directory is a fixed field header followed by variable data fields. The fixed fields give the address to, and the length of the variable fields. The items of interest in the external data base are selected, and fields in the ABACUS format are allocated by the reformatting program. Depending on the amount of information on the external tape, the identification process differs from one format to another.

Among the more extensive format in the databases are ERIC Report Resume Master Data Set and Government Reports Announcements many of which fields are not applicable in the shorter format of databases containing references to journal articles. Not all fields in the different databases are of interest to the users. Thus, at present, some fields are deleted when reformatting into the ABACUS. Table 1-2 shows the ERIC Report Resume Master Data Set Fields and the International Food Information Service - IFIS - and their treatment in the ABACUS record. Even if documentation is provided by a data base producer, the reformatting specification is written after inspection of tape dumps.

In general, the reformatting of the different tape formats is rather straightforward work of 30 hrs programming, even if they deviate from the ISO interchange format into which, it is hoped, they will eventually change. Essentially, the allocation of fields in the ABACUS program depends on the fields identification numbers within the record types for reports and articles. As can be seen from Table 1-2 the 26 fields in the ERIC format yield 5 fields and the 17 fields in the IFIS yields 8 fields in the ABACUS set of searchable fields. The search terms can operate within these, since they are specified with regard to the type of field in which they are to be searched.

The Reformatting of ERIC Report Resume Master Data Set Fields into the ABACUS Format

ERIC		ABACUS		
Field name	Field identification no. in hexadecimal	Searchable	Printout	Deletion
Sequence	0000			x
Add Date	0001			x
Change Date	0002			x
Accession Number	0010		x	
Clearinghouse Accession Number	0011		x	
*Other Accession No.	0012			x
*Program Area	0014			x
*Publication Date	0017		x	
Title	001A	x	x	
Personal Author	001B	x	x	
*Institution Code	001C			x
*Sponsoring Agency Code	0020			x
Descriptor	0023	x		
Identifier	0024			x
*EDRS Price	0025			x
*Descriptive Note	0026			x
Issue	002B		x	
Abstract	002C			x
*Report Number	002D		x	
*Contract Number	002E			x
*Grant Number	002F			x
*Bureau Number	0030			x
*Availability	0031			x
Journal Citation	0032	x	x	
*Institution Name	0080	x		
Sponsoring Agency Name	0084			x

* Not Used in CIJE

The reformatting of IFIS data set fields into the ABACUS format

FSTA	ABACUS			
Field name	Field identification no.	Searchable	Printout	Deletion
Year, vol., no., category, running no. of printed abstract	010		x	
Authors	030	x	x	
Author annotation	035			x
Year	036		x	
Title in English	040	x	x	
Original title if not in English	041			x
Title annotation	042		x	
Journal name, patent country	050	x	x	
Vol., issue, page, patent no.	055		x	
Number of cited references	056		x	
Language	057	x	x	
Affiliation	058	x		
Abstract	080			x
Initial of abstractor	081			x
Heading	101	x		
.	.	x		
Heading	.	x		

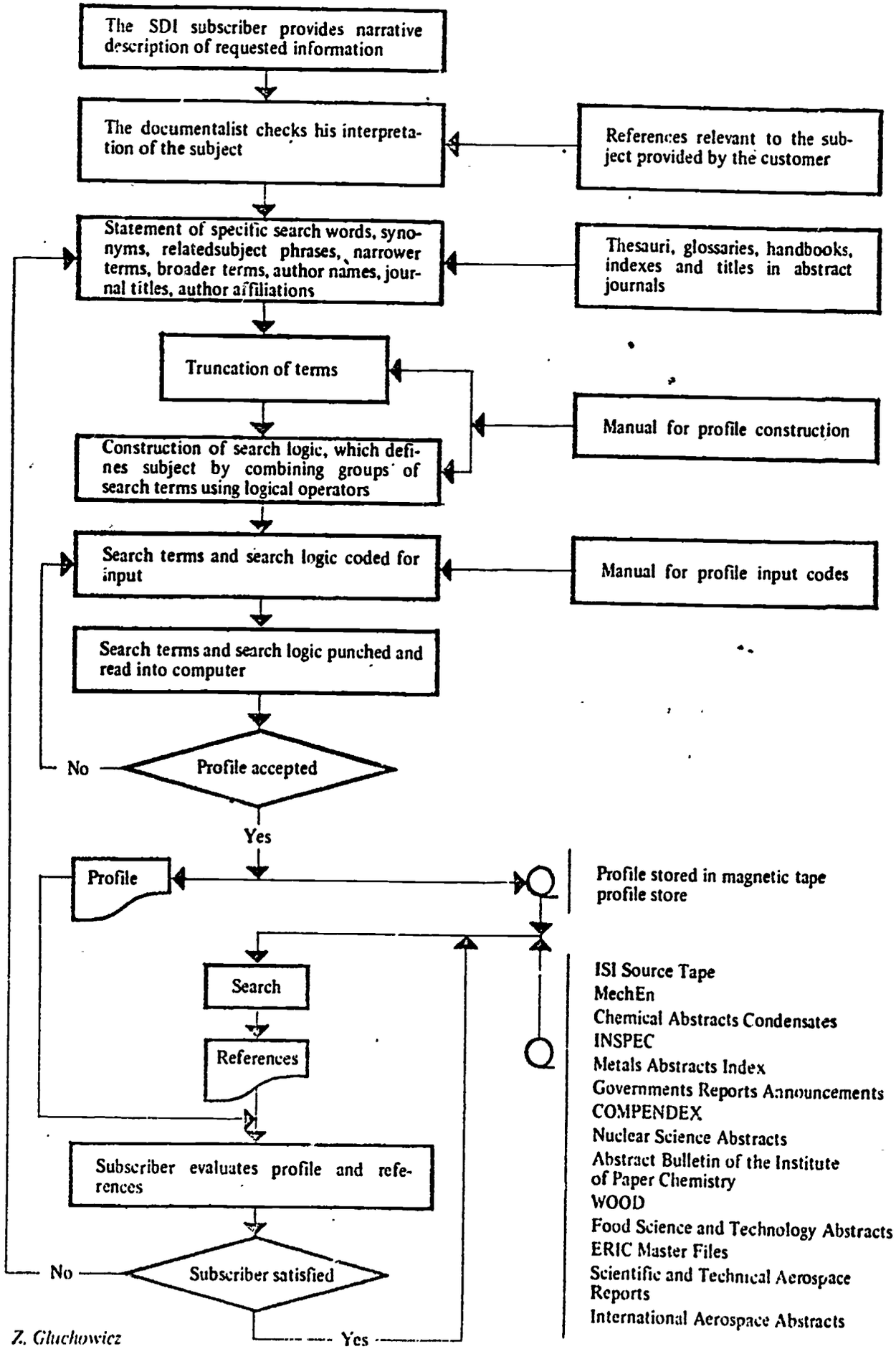
6. PROFILE CHARACTERISTICS

The construction and revision of query profiles is an essential task in an SDI system which demands an effort both from the user and the subject specialist. When a user wants to submit a question to the SDI system he is requested to formulate his field of interest in natural language, which means in a normal narrative way, describing his interest in some detail. It has proved very useful for the user also to supply some references to papers which he considers relevant to his query. He could also provide a list of significant terms and, if possible, make a draft of the actual search profile. The staff has prepared a Profile Design Manual which explains the principles of a computer-operated information retrieval system and describes all details of the profile construction.

The interaction between the staff and the user is essential for a successful search. On the basis of the user's statements the subject specialist specifies the question by making a list of significant terms, which might occur as potential words in the titles of documents. Among the staff there are subject specialists in education, psychology, business administration, electrical & mechanical engineering, chemistry, physics, etc. Furthermore, the list might also include authors, affiliations, and journal titles. As the system permits search both on keywords and on natural language used in titles, the subject specialist uses thesauri, handbooks, dictionaries, and all other means he might find helpful and relevant for the formulation of the profile. He has to make a special point of checking the printed volumes of the corresponding databases to find the occurence of terms when used alone or in combination with other terms. A generalized flow chart, Fig. 1. has been constructed by Zofia Gluchowicz (2).

While the keywords must be written exactly as they appear in the Thesaurus and on the tape, the free text terms in potential titles can be truncated both at the beginning and at the end. Truncation facilitates retrieval of items containing word fragments which are common to different forms of a word, and words within words can be searched for. As will be seen from examples below, suffix (right-hand) truncation occurs very often, while prefix (left-hand) truncation is more unusual. Both suffix and prefix truncation is, on the other hand, more common. For example, the truncated term /CASSETT/, where the slashes stand for truncations, will retrieve STEREOCASSETTES, VIDEOCASSETTE, CASSETTE-RECORDER, CASSETTE/CARTRIDGE, etc.

GENERALIZED FLOW CHART FOR PROFILE CONSTRUCTION



Z. Gluchowicz

As can be seen from Fig. 2-3 the terms are numbered sequentially in the profile printout to facilitate updating. The terms are also grouped together, and the groups are indicated by capital letters A,B,C etc. Terms, or groups of terms, are linked together in a logical manner by using "and", "or", and "not" logic. The number of terms in one profile might be up to the system-allowed 150 in ABACUS. In the new VIRA program there are no such restrictions. On the other hand, as charging policy is to count 30 terms as one profile, the average number of terms per profile varies around 24.

The printout of the profile also includes a description in natural language of the query, the search logic, and the list of terms classified according to type of terms such as words, keywords, author names etc. The profile printout and every updating of it is sent to the user. For verification a copy of the profile as well as a copy of the search results are kept in the files of the service, transferred every 9 months into microfilm cassettes.

The user's responses to early selections based on the first profile approximation to his field of interest are used for improving the profile. Thus, the maintenance of the profile is carried out by adding new terms, and subtracting old ones which do not give satisfactory results, or by opening and tightening the logic. False co-ordinations between search terms from different term groups can also be detected and should be avoided. While constructing the initial profile we try to choose the logical strategy considering the user's wishes, and accordingly decide on the degree of restrictivity for the initial computer run. Often we use a less restrictive logic, i.e. not too many "and" or "not" restrictions, in the initial profile, even if it will result in an output of many irrelevant references, i.e. noise, and then, after a few searches adjust the profile on the basis of the user's evaluation of the output.

Figure 2.

Profile 70E
Subject: Audiovisual aids for the mentally retarded.
Data bases: ERIC, ISI, INSPEC.
Logic: A & B

Term No.	Term Group	Search terms	Weight	Term Type
010	A	TAPE RECORD/	2	KEYWORD
020	A	VIDEO TAPE RECORD/	2	KEYWORD
030	A	EDUCATIONAL TELEVISION/	2	KEYWORD
040	A	INSTRUCTIONAL TELEVISION/	2	KEYWORD
050	A	AUDIOVISUAL/	10	KEYWORD
060	A	CASSETT/	2	WORD
070	A	CARTRIDGE/	2	WORD
080	A	EVR	2	WORD
090	A	VTR	2	WORD
100	A	VCR	2	WORD
110	A	ETV	2	WORD
120	A	ITV	2	WORD
130	A	CTV	2	WORD
140	A	SELECTAVISION/	2	WORD
150	A	TELEVISION	2	WORD
160	A	TV	2	WORD
170	A	/VIDEO/	2	WORD
180	A	CARTRIVISION/	2	WORD
190	A	8MM/	2	WORD
200	A	AUDIOVISUAL/	10	WORD
210	A	AV	10	WORD
220	A	A-V	10	WORD
230	A	VIDICORD/	10	WORD
240	A	VISUAL AID/	10	WORD
250	A	MEDIA/	2	WORD
260	A	PICTURE/	2	WORD
270	A	LONG-DISTANC/	10	WORD
280	A	AUDIO-VISUAL	10	WORD
290	B	EDUCATIONALLY DISADVANTAG/	2	KEYWORD
300	B	LOW ABILIT/	2	KEYWORD
310	B	SLOW LEARNER/	2	KEYWORD
320	B	MENTALLY HANDICAP/	10	KEYWORD
330	B	EDUCABLE MENTALLY HANDICA/	10	KEYWORD
340	B	RETARDED/	10	KEYWORD
350	B	RETARDATION/	10	KEYWORD
360	B	MENTAL RETARDATION/	10	KEYWORD
370	B	EXCEPTIONAL/	2	KEYWORD
380	B	SPECIAL/	2	KEYWORD
390	B	RETARD/	10	WORD
400	B	LOW/	2	WORD
410	B	SLOW/	2	WORD
420	B	FAILUR/	2	WORD
430	B	DISADVANTAG/	2	WORD
440	B	HANDICAP/	2	WORD
450	B	BELOW/	10	WORD
460	B	EXCEPTION/	2	WORD
470	B	DROPOUT/	2	WORD

In total 45 searchwords, of which 14 are keywords from the ERIC Thesaurus.

Profile 26U

Subject: Electronic circuits and systems

Data bases: INSPEC

Logic: A +B*(C +D +E) +C*(E +F +G +H +K +L +M) +
L*M +E*(K +M) +G*N +P*(H +R +B*K) -S

Term No.	Term Group	Search terms	Weight	Term Type
0008	A	* AUTOMATA THEORY*	2	WORD
0009	A	* COMPUTER DESIGN*	2	WORD
0010	A	* DIGITAL SYSTEM*	2	WORD
0011	A	* LOGIC SYSTEM*	6	WORD
0012	A	* MACHINE LOGIC*	2	WORD
0013	A	* SEQUENTIAL MACHINE*	6	WORD
0014	A	* SYNCHRONOUS SYSTEM*	2	WORD
0015	B	* NETWORK*	6	WORD
0016	C	* LOGIC*	6	WORD
0017	D	* DIGITAL*	2	WORD
0018	E	* SEQUENTIAL*	2	WORD
0019	F	* ALGORITHM*	2	WORD
0020	F	* AUTOMAT*	2	WORD
0021	F	* COMBINAT*	2	WORD
0022	F	* PARTITION*	2	WORD
0023	G	* FUNCTION*	2	WORD
0024	H	* SIMULAT*	2	WORD
0025	K	* SYNTHESIS*	2	WORD
0026	L	* LANGUAGE*	2	WORD
0027	M	* DESIGN*	2	WORD
0028	N	* MULTIPLE OUTPUT*	2	WORD
0029	N	* MULTI-VALUE*	2	WORD
0030	P	* B03*	2	CLASSIFICATION CODE
0031	P	* B046*	2	CLASSIFICATION CODE
0032	P	* C90*	2	CLASSIFICATION CODE
0033	P	* C92*	2	CLASSIFICATION CODE
0034	P	* C93*	2	CLASSIFICATION CODE
0035	R	* NAND *	2	WORD
0036	R	* NOR *	2	WORD
0037	R	* FLIPFLOP*	2	WORD
0038	R	* FLIP FLOP*	2	WORD
0039	R	* MINIMI*	2	WORD
0040	S	* FILTER*	98	WORD

7. PROCESSING METHODS AND COSTS

An inevitable characteristic of large retrieval systems is, that a strategy for searching a small or medium size data base might differ significantly from a search strategy for a large base. During the five years our search methods have passed through the mere masking-off technique, yielding search times proportional to the number of references and terms in the profiles, into a more elaborate technique making use of hashcoding and tree structure searches, thus arriving at an almost logarithmic increase in time when the number of terms in the profile grow. The newest program, having the acronym VIRA and written by Rolf Larsson, is run in parallel with ABACUS (Zennaki 3) The present profile program, PROSA, includes 2,500 statements in COBOL, and the VIRA search program counts 2,000 statements in IBM assembler language.

In order to carry out a rough check of the performance of the profiles on a "management by exception" basis, two statistical tools have been developed. The critical values of the printout to a user are (1) an abundance of references, and (2) no printout. In order to reveal these extremes, every search results in search statistics indicating the number of references for each profile. The form is designed like the scale of the speedometer of many cars, the longer the row of "stars" the more the reason to put ones foot on the brake. Fig. 4 displays part of the search statistics for a run on ERIC. The columns give the number of references to the first digit, the second, etc. Thus,the first profile has resulted in 6+40 = 46 references, the second in 8+60+300 = 368 references. On the other hand, profile No 26R has given no output. Furthermore, at the bottom on the form an indication is given of which profiles have received no hits, and those which have received more than 40 hits.

These search statistics give an indication of where the exceptional cases are located among the profiles. The next step is to analyse what causes the no-hits or the great number of hits. In order to find out about the latter case, a listing is also given for every profile stating which terms or term combinations have caused the printout including the frequencies of these terms.See Fig. 5 in which case the first step would be to analyse the combination MEASUREMENT TECHNIQUES and MEASUREMENT INSTRUMENTS which occurs 13 times, perhaps in order to change the logic or to place these words in separate groups, if they have given rise to many irrelevant references. The second column in Fig. 5 indicates the weights we are experimenting with which will be discussed later on.

8 SEARCHING KEYWORDS AND WORDS IN TITLES

The ABACUS program is designed in such a way that it can process natural language by searching titles and/or abstracts. In the case of another data base, Science Citation Index Source Tapes, the ISI tapes, which covers 2,000 journals there are no keywords or other subject indicators than the titles. Thus, free text search is the only way to open the files. Free text search can be regarded as using a set of skeleton keys to open up any machine readable file. Some files make use of keywords chosen from a corresponding thesaurus of descriptors. Searching these keywords become an additional means for the subject specialist or the user to augment the search performance of the files containing keywords compared with the ISI tapes. When a data base contains keywords, we have recommended that they should be used in combination with words in natural language. In a multi-data base environment the same profile in natural language can easily be used on various data bases, while the use of keywords is restricted to each specific data base which has to be taken into account when formulating the profile. Many of our profiles are searched on several databases since our main principle is to answer the query in its broadest sense disregarding from which data base the responding references will stem.

Especially for questions of inter-disciplinary nature it is obvious that they should be processed on several data bases in order to assure good coverage. It is true, however, that the reformulation of a query into a profile for the SDI system takes place in a kind of dialogue with the computer, focusing on one data base at a time considering both the terminology used in free text, and the metalanguage of keywords or other subject indicators. In order to arrive at a standardization of the query formulation, allowing for different degrees of complexity of natural text and metalanguages, a method has been developed for translation between the various scientific disciplines reflected in the data bases by the generation of vocabularies and concordance for words in natural language and the various thesauri used.

We have started work in this area by the compilation of word frequency lists for various data bases as ERIC, CAC, INIS, and ISI.

That the use of the language (the scientific "jargon") is different in various disciplines has been displayed when compiling frequency lists for these disciplines. So, for instance, was the first significant word in the INIS system - nuclear energy - REACTOR, and the first in CAC - organic chemistry - ACID, in ERIC - EDUCATIONAL. The non-informative

Search statistics from a run on ERIC tapes

0131.xxxxxx	.xxx	.	.
02E1.xxxxxxxx	.xxxxxx	.xxx	.
02U1.xxx	.x	.	.
07F1.	.xx	.	.
08F1.xxx	.	.	.
10A1.xx	.xxxxxxxx	.xxxx	.
11E1.	.xxxxxxxx	.x	.
12F1.x	.xxxx	.	.
13F1.xx	.xxxx	.xx	.
14F1.xxxxx	.x	.x	.
15F1.xxxxxxxx	.	.	.
16R1.xx	.x	.	.
19C1.xxxxxxx	.xxxxxx	.	.
26R1.	.	.	.
28E1.xxxx	.xxxxxxxx	.	.
31C1.xxx	.	.	.
3141.xxxxxx	.xx	.	.
32S1.xxxxxxxx	.xx	.	.
36D1.x	.x	.xx	.
36F1.xxxxxxxx	.xxx	.	.
38E1.	.xxxx	.xx	.
38F1.xxxxxxxxx	.x	.	.
39F1.xxxxx	.xxx	.	.
40G1.xxxxxx	.xx	.	.
41F1.xxx	.xxxxxxxx	.	.
44G1.	.	.	.
45B1.xxxx	.	.	.
45D1.xx	.xxxxx	.	.
45F1.xxxxxxxx	.x	.xxx	.
45G1.xxxxxxx	.xxxxx	.	.
51A1.xxxxxxx	.	.xx	.
51F1.xxx	.	.xx	.
52F1.x	.x	.	.
54F1.xxxxx	.xxxxxxxx	.	.
5411.xxxxxxx	.xxxxxxxxx	.	.
56E1.xx	.xxxx	.	.
56F1.xxxx	.xx	.	.
563.xxxxxx	.xxx	.x	.
57E1.	.xxx	.	.
58E1.xx	.xxx	.	.
58R1.xxxxxxxxx	.xxxxxxx	.	.
59E1.xxxxx	.x	.	.
70E1.xxxxxxxx	.xxxx		

Following profiles gave no output
26R1 44G1 6241 70F1 80A1 8851

Following profiles gave more than 40 hits
0131 02E1 10A1 11E1 12F1 13F1 14F1 19C1 28E1 36D1 38E1
41F1 45D1 45F1 45G1 51A1 51F1 54F1 5411 56E1 5631 58R1
70E1

Frequencies of coincidences of profile 64G.

1	VIKT=30,00	* BEHAVIOR* CLASSROOM OBSERVATION TECHNIQUE*
2	VIKT=30,00	* BEHAVIOR* CLASSROOM* CLASSROOM OBSERVATION T
1	VIKT=30,00	* BEHAVIOR* PERFORM* CLASSROOM* TEACH* CLASSRO
1	VIKT=30,00	* BEHAVIOR* PUPIL* CLASSROOM OBSERVATION TECHN
3	VIKT=30,00	* BEHAVIOR* TEACH* CLASSROOM OBSERVATION TECHN
12	VIKT=30,00	* CLASSROOM OBSERVATION TECHNIQUE*
5	VIKT=30,00	* CLASSROOM* CLASSROOM OBSERVATION TECHNIQUE*
1	VIKT=30,00	* CLASSROOM* TEACH* CLASSROOM OBSERVATION TECH
2	VIKT=30,00	* EDUCATION* CLASSROOM OBSERVATION TECHNIQUE*
2	VIKT=30,00	* OBSERVATION* CLASSROOM OBSERVATION TECHNIQUE
2	VIKT=30,00	* PUPIL* TEACH* CLASSROOM OBSERVATION TECHNIQU
1	VIKT=30,00	* TEACH* BEHAVIOR* CLASSROOM OBSERVATION TECHN
5	VIKT=30,00	* TEACH* CLASSROOM OBSERVATION TECHNIQUE*
3	VIKT=30,00	* TEACH* METHOD* CLASSROOM OBSERVATION TECHNIQ
1	VIKT=30,00	* TECHNIQUE* CLASSROOM OBSERVATION TECHNIQUE*
1	VIKT=32,00	* OBSERVATION* TEACH* METHOD*
1	VIKT=35,00	* OBSERVATION* EDUCATION* EVALUATION TECHNIQUE
1	VIKT=41,00	* OBSERVATION* BEHAVIOR* MEASUREMENT TECHNIQUE
2	VIKT=50,00	* ACHIEVEMENT* MEASUREMENT TECHNIQUES * MEASUR
1	VIKT=50,00	* ADUCATION* TEACH* TECHNIQUE* MEASUREMENT TEC
1	VIKT=50,00	* INSTRUMENT* MEASUREMENT TECHNIQUES * MEASURE
13	VIKT=50,00	* MEASUREMENT TECHNIQUES * MEASUREMENT INSTRUM
1	VIKT=50,00	* TEACH* MEASUREMENT TECHNIQUES * MEASUREMENT
1	VIKT=50,00	* TECHNIQUE* MEASUREMENT TECHNIQUES * MEASUREM
1	VIKT=52,00	* ACHIEVEMENT* MEASUREMENT TECHNIQUES * EVALUA
1	VIKT=52,00	* CLASSROOM* MEASUREMENT INSTRUMENTS * EVALUAT
1	VIKT=52,00	* INSTRUMENT* MEASUREMENT INSTRUMENTS * EVALUA
1	VIKT=52,00	* INSTRUMENT* MEASUREMENT TECHNIQUES * EVALUAT
3	VIKT=52,00	* MEASUREMENT INSTRUMENTS * EVALUATION TECHNIQ
4	VIKT=52,00	* MEASUREMENT TECHNIQUES * EVALUATION TECHNIQU
1	VIKT=52,00	* STUDENT* MEASUREMENT TECHNIQUES * EVALUATION
1	VIKT=52,00	* TEACH* TECHNIQUE* MEASUREMENT INSTRUMENTS *
2	VIKT=57,00	* EVALUATION TECHNIQUES * CLASSROOM OBSERVATIO
1	VIKT=57,00	* TEACH* EVALUATION TECHNIQUES * CLASSROOM OBS
1	VIKT=57,00	* TEACH* STUDENT* BEHAVIOR* EVALUATION TECHNIQ
1	VIKT=60,00	* OBSERVATION* TEACH* BEHAVIOR* CLASSROOM*
1	VIKT=70,00	* OBSERVATION* TEACH* STUDENT* CLASSROOM OBSER
1	VIKT=80,00	* OBSERVATION* MEASUREMENT TECHNIQUES * MEASUR
1	VIKT=82,00	* CLASSROOM* MEASUREMENT INSTRUMENTS * EVALUAT
1	VIKT=86,00	* OBSERVATION* CLASSROOM* TEACH* METHOD* CLASS

8 SEARCHING KEYWORDS AND WORDS IN TITLES

The ABACUS program is designed in such a way that it can process natural language by searching titles and/or abstracts. In the case of another data base, Science Citation Index Source Tapes, the ISI tapes, which covers 2,000 journals there are no keywords or other subject indicators than the titles. Thus, free text search is the only way to open the files. Free text search can be regarded as using a set of skeleton keys to open up any machine readable file. Some files make use of keywords chosen from a corresponding thesaurus of descriptors. Searching these keywords become an additional means for the subject specialist or the user to augment the search performance of the files containing keywords compared with the ISI tapes. When a data base contains keywords, we have recommended that they should be used in combination with words in natural language. In a multi-data base environment the same profile in natural language can easily be used on various data bases, while the use of keywords is restricted to each specific data base which has to be taken into account when formulating the profile. Many of our profiles are searched on several databases since our main principle is to answer the query in its broadest sense disregarding from which data base the responding references will stem.

Especially for questions of inter-disciplinary nature it is obvious that they should be processed on several data bases in order to assure good coverage. It is true, however, that the reformulation of a query into a profile for the SDI system takes place in a kind of dialogue with the computer, focusing on one data base at a time considering both the terminology used in free text, and the metalanguage of keywords or other subject indicators. In order to arrive at a standardization of the query formulation, allowing for different degrees of complexity of natural text and metalanguages, a method has been developed for translation between the various scientific disciplines reflected in the data bases by the generation of vocabularies and concordance for words in natural language and the various thesauri used.

We have started work in this area by the compilation of word frequency lists for various data bases as ERIC, CAC, INIS, and ISI.

That the use of the language (the scientific "jargon") is different in various disciplines has been displayed when compiling frequency lists for these disciplines. So, for instance, was the first significant word in the INIS system - nuclear energy - REACTOR, and the first in CAC - organic chemistry - ACID, in ERIC - EDUCATIONAL. The non-informative

words as FOR and TO occur in almost the same order in these data bases. The following remarks based upon our experience might illuminate the efficiency of descriptors in a thesaurus. The combined search strategy we use, mixing keywords and words in free text, reveals that the present indexing habit in some data bases of using keywords identical to words in the titles is futile. If some of the keywords instead took the place of broad subject categories it would add a new dimension to the search. This is, for instance, the case with the data base INSPEC.

A study should also be made about the proportion of titles that are not useful as content indicators and, thus, not suitable for free text searching. If only a small amount of titles are meaningless, a human indexing using thesaurus keywords should be questionned.
On the other hand, if something needs to be done, especially if we believe that keyword indexing is necessary for the quality of printed indexes or for future on-line retrieval systems of the RECON type, title augmentation of automated keyword assignement seem to be attractive alternatives to expensive human indexing. Such a strategy might cause authors to improve the information content of their titles. This has happened in areas where KWIC indexing technique is used.
Because of the costs of indexing we could never afford it for our own data base in mechanical engineering, wood, paper and pulp industry, covering 250 journals (60,000 references/yr) in three languages. Only title augmentation is permitted in case of short titles (less than 60 characters). We know that we can give satisfaction to the users by free text searching only, because at present, we receive orders for several hundreds photocopies a month as a result of the output.

9. EVALUATION AND FEED-BACK

At present 1100 users receive SDI service on our databases. After five years of operation on tapes in general we feel that we are still just scratching the surface of computerized information retrieval. We think, for instance, that the printout we now deliver as answers to the queries should go through further refinement before reaching the user. When we consider the construction of a profile as reflecting a specific query, it is difficult to provide a measure of its effectiveness,especially as our practice is to retrieve references from multiple files. Questions about recall and precision lose interest. The essential measure which we can assess is the user's satisfaction which can be expressed on a scale from highly relevant to irrelevant, or by counting the number of documents he orders.

Time and costs of the computer are other factors which can be measured, between computer costs plus the costs for the tapes and the subscription fee for the profiles, leaving other costs, e.g. the construction of the profiles to be defined as common library costs.

The delay time for the same reference appearing in the various services has been studied. We know that ISI is much faster than COMPENDEX or INSPEC, and also than ERIC. However, delay time often does not have a significant effect on the user. It happens instead when he receives an early reference that he judges it as of low interest or irrelevant, while the same reference appearing 3-6 months later is evaluated as very interesting, and he orders a copy. In several cases, it seams as the continuous SDI service has a sort of learning effect on the user.

10. METHODS TO ESTABLISH A HELPFUL OUTPUT ORDERING

This paper is not intended as a primer on information retrieval but the reader might already have noticed in Fig. 2 ,3 and 5 that there are indications of a weighting procedure (VIKT = WEIGHT). We should, therefore, like to mention that we are experimenting with various weighting methods in order to establish a helpful ordering of the output so that references early on the list should have higher probability of interest to the individual user than the later ones.

The method shown in Fig. 2 and 3 is based upon the assumption that the words used in the profile and the words occuring in a reference are related in such a way that the more the words co-occur, the higher the probability that the reference is relevant to the query.

This gives us one way of ordering the output. Thus, we note the number of co-occurences and let the search logic operate arithmetically to

LITTERATURLISTA

Formulär 5

datum
18/02/72

beställare

kontaktperson

företag/institution

adress

postadress

sökprofil nr
70E1

FILM PRODUCTION WITH THE MENTALLY RETARDED
JUNKALA, JOHN
EDUC TRAINING MENT RETARDED; 4; 2; 75-9
69 APR Education & training of the Mentally Retarded EJ006225
VIKT=200,00 * AUDIOVISUAL* MENTALLY HANDICAPPED* RETARD*

A VOCATIONAL PICTURE INTEREST INVENTORY FOR EDUCABLE RETARDED YOUTH
BECKER, RALPH L. FERGUSON, ROY E
EXCEPT CHILDREN; 35; 7; 562-3
69 MAR EJ002512
VIKT=200,00 * PICTURE* MENTALLY HANDICAPPED* RETARD*

THE USE OF FILMSTRIPS IN TEACHING PERSONAL HYGIENE TO THE MODERATELY RETARDED ADOLESCENT
THOMPSON, MARY MARTHA FAIBISH, GEORGE M
EDUC TRAINING MENT RETARDED; 5; 3; 113-8
OCT '70 EJ026874
VIKT=200,00 * AUDIOVISUAL* MENTALLY HANDICAPPED* RETARD*

VIDEOTAPE AS A TEACHING TOOL
AUSTIN, JAMES T.
EXCEPT CHILDREN; 35; 7; 557-8
69 MAR EJ002510
VIKT=120,00 * AUDIOVISUAL*VIDEO* MENTALLY HANDICAPPED*

USE OF A LISTENING STATION FOR INTRODUCTION A READING LESSON TO DISADVANTAGED EMR JUNIOR HIGH SCHOOL STUDENTS
YATES, JOHN R., JR.
EDUC TRAIN MENTING RETARD; 4; 1; 29-31
69 FEB EJ000803
VIKT=120,00 * AUDIOVISUAL* MENTALLY HANDICAPPED* DISADVANT

MEDIA SYSTEMS AND THE HANDICAPPED CHILD
MCINTYRE, KENNETH
AUDIOVISUAL INSTR; 14; 9; 21-3
69 NOV EJ010851
VIKT=120,00 * AUDIOVISUAL* MEDIA* HANDICAP*

THE INSTRUCTIONAL MATERIALS CENTER NETWORK FOR HANDICAPPED CHILDREN AND YOUTH
ERICKSON, DON
AUDIOVISUAL INSTR; 14; 9; 41
69 NOV EJ010855
VIKT=100,00 * AUDIOVISUAL* HANDICAP*

datum
72-08-29

MEASUREMENTS OF AURORAL ELECTRIC-FIELDS WITH A DIFFERENTIAL RETARDING POTENTIAL ANALYZER
CARLSON CW.
J AN GEOPHY 52 329 74 8 44 12817
DIFFERENT ANALY

beställare kontaktperson

företag/institution

adress

postadress

sökprofil nr
26UU INSPEC

VIRTUAL MEASURES FOR COMPUTER SIMULATION EXPERIMENTS OPERATIONS RESEARCH SOCIETY OF AMERICA 41ST ANNUAL MEETING. ABSTRACTS ONLY
CARTER, G. IGNALL, E.
BULL. OP. RES. SOC. AM. (USA)26-28 APR 1972
VOL.20, SUPPL. 1 B-175 SPRING 1972
VIKT=4,00 * SIMULAT* C90* 12 F

SIMULATION OF ANALOGUE CIRCUITS: THE IMAG II PROGRAM
ARNOULD, J. LEFAOU, C.
ELECTRON. AND MICROELECTRON. IND. (FRANCE)
NO.153 45-9 15 MARCH 1972 7
FRENCH
VIKT=4,00 * SIMULAT* B03* 13 F

OPTIMIZATION OF FAULT SIMULATION BY PREANALYSIS
VOGELSBERG, R.E.
IBM TECH. DISCLOSURE BULL. (USA)
VOL. 14, NO. 8 2508-9 JAN. 1972
VIKT=4,00 * SIMULAT* C92* 14 F

IDENTIFYING SOURCES OF UNKNOWN LEVELS GENERATED DURING THREE-VALUE FAULT SIMULATION
VOGELSBERG, R.E.
IBM TECH. DISCLOSURE BULL. (USA)
VOL. 14, NO. 8 2510-12 JAN. 1972
VIKT=4,00 * SIMULAT* C92* 15 F

AN ANALYSIS OF TRANSPORT DELAY SIMULATION METHODS
KNOWLES, J.B. LEGGETT, D.W.
RADIO AND ELECTRON. ENG. (GB)
VOL.42, NO.4 172-8 APRIL 1972 15
VIKT=4,00 * SIMULAT* B03* 16 F

A TRANSPORT DELAY SIMULATOR USING DIGITAL TECHNIQUES
KEATS, A.B. LEGGETT, D.W.
RADIO AND ELECTRON. ENG. (GB)
VOL.42, NO.4 179-84 APRIL 1972 2
VIKT=4,00 * DIGITAL* SIMULAT* B046* 17 F

arrive at the values upon which we base the orders. As can be noted from the profile 70E in Fig. 2, the weight 2 in general is assigned to all terms. However, the user has regarded some terms of greater importance and assigned the weight 10 to them. The three words which pick up the first reference in the printout in Fig. 6 have all the weight of 10, two of which are in the same term group, thus, 10 + 10. The logical Boolean operation "and" is translated into multiplication, so the complete expression will be: 10 x (10+10) = 200, as the weight shows. To the four words which pick up the first reference in the printout in Fig. 7 the following weights have been attached in the profile, see Fig. 3, NETWORK-6, LOGIC-6, C 92-2, NAND-2. According to the search strategy of this profile the reference becomes the weight 6x6+2x2=40. In this case it seams to have worked to the user's satisfaction, since he has ordered a copy by circling the reference. Usually we do not influence the user to put in subjectively assigned weights, as we should like to find out more about the objectively assigned weights. This brings us back to the list of word frequencies dealt with under Chap.7. We could order the references based upon the frequencies of the words in the data base which is our next step in preparation. The underlying reasoning is as follows.

When forming the logical expression in a keyword based system arranged as an inverted file, it is common to base the logical expression upon the number of documents pinned to each keyword. This number indicates the frequency with which this keyword has been used for indexing. Thus, on-line searches on a display terminal usually end by forming the logical expression that gives the minimum output. This means that high frequency terms are looked upon as having less value than those with low frequencies. In a free text search system in the batch processing mode, a search can be based also upon term frequencies using natural language if we build a frequency table from a large sample of references of each data base, say around 30,000 references. The values for ordering could then be established as the sum of the values of the co-occurring terms, if those are expressed as 1/n, where n is the frequency of the term given by the frequency table (Tell 4). Such frequency tables are under construction for several data bases.

The weighting procedure is only the first step. We are going to study parsing and computational linguistic methods in order to find out the contribution such methods can give to the output ordering. We hope to arrive at shorter lists by introducing a cut-off when the weights are too low, thus saving computer and user time.

11. PERSONNEL AND TRAINING

Being responsible for exploring the utility of computerized information services to scientific research, higher education and industry, we have felt that one task has been to carry out research and development of the kind which has been disclosed above. The other tasks are production, management, clerical support, and supporting library service. The overall staff picture for running the SDI service is 12 full-time equivalents. The number of subject specialists are 8, clerical equivalents 4, and programmers 1. In the transitory state we are at present, operating with two systems, ABACUS and VIRA, the profile updating is laborious which has made it difficult, for example, to devote time to the construction of group profiles of interest in several areas. SDI is tailor-made for the individual and requires personal attention of the subject specialist, and becomes relative time-consuming, while group profiles are cheaper in updating without the necessity to adapt to individual requirements.

Also the library back-up service has been put under pressure since the introduction of the SDI service. Even if requests for copies of the references put out of some files are shifted over to other libraries where some microfiche collections are located, most references to journal articles and technical reports are handled by our library from its collections or by inter-library loans. In many cases photocopies are ordered from the National Lending Library in Boston Spa, U.K. This follow-up service is found to be important in order to keep the interest of the users.

The effectiveness of the search profile is, to a high degree, dependent on the active interest of the subscriber. The user is more able to influence the effectiveness of his search profile if he knows the basic principles of the computer-operated information retrieval system and profile construction technique. Therefore we have organized one-, two- and ten-days educational seminars with lectures and exercises in profile construction, see Table 3-5. Research engineers, production engineers and draftsmen of different levels have participated in these seminars. All of them had encountered the increasing need for up-to-date information in their daily work. The participants were not only informed about the principles of the SDI system, but were also given an introduction to manual information retrieval methods, see Table 5. This was done because the initial intellectual effort placed on the the user, when he has to define his problem, is the same for both methods of information retrieval.

Seminar on the structure and use of scientific and technical literature for scientists, engineers, and technicians.

Programme

Day 1.	Morning	Introduction to seminar. Tour of the library.
		Structure of scientific and technical literature.
		Guides to primary and secondary information sources.
	Afternoon	The technique of literature search by conventional methodes.
		Practical work: Training in the use of scientific literature. Participants perform literature search on specially chosen items.
		Discussion of seminar.
Day 2.	Morning	Special libraries, information centres, documentation services.
		Computerized information retrieval: The SDI system at the Institute, profile performance and users' feedback.
	Afternoon	Practical work: Participants perform profiles on chosen items.
		Discussion of seminar.

Zofia Gluchowicz

Seminar on the SDI system at the Royal Institute of Technology
(Selective Dissemination of Information)

	Programme	
Day 1.	Morning	Introduction to seminar.
		SDI from the user co-ordinator's viewpoint. Description of data bases, profile performance, feedback, evaluation, profile adjusting.
		SDI from the users' viewpoint. SDI users relate their experience of the SDI service.
	Afternoon	Practical work: Training in profile performance on items chosen by the participants.
Day 2.	Morning	SDI from the system designer's and the programmer's viewpoints.
		Practical work continued as above.
	Afternoon	Development trends and future prospects of computerized information retrieval.
		Discussion of seminar.

Zofia Gluchowicz

One day seminar on the SDI system at the Royal Institute of Technology (Selective Dissemination of Information)

Programme

Morning	Introduction to seminar
	Presentation of tapes service and subject categories covered.
	Profile construction for SDI service, evaluation, feedback.
Afternoon	Practical work: Participants perform individual search profiles for searching on the different tapes.
	Discussion on seminar.

About 70 engineers and scientists participated in the seminars.

Zofia Gluchowicz

The user will more easily associate the new technique with the traditional methods and he will be better aware of what the SDI service can offer regarding literature coverage and timeliness. In this way the interest for the SDI service has been intensified and the user takes more active part in the handling of the profiles. These seminars are much appreciated and they are given in different parts of Sweden. Lectures on and training in profile construction have also been included in the curriculum for the fourth year for the students of the Institute. The courses have been given by the library staff.

During the two-months course in information and documentation techniques for graduates in science and technology, 60 hours were reserved for lectures and training in computerized documentation and profile construction.

Our experience from trying to market the data bases to scientists and people in industry has been that the most effective means is one-day seminars where afternoon sessions is devoted to group work when every participant under the guidance of one of our staff constructs a profile in his field of interest, see Table 5. We promise then to run it on a trial basis free of charge for a few months. Such a procedure of "taking the service to the user", has appeared successful in attracting potential users.

12. THE ON-LINE INTERACTIVE MODE

We have now arrived to the stage when, as information centre, we have started to use terminal equipment for on-line access to computer stored information in big information data banks. The salient component in this man-machine interactive system is the remote console. In our case it is a portable input/output terminal which generates and displays information on a standard television receiver, accepts information from a keyboard and communicates with the computer which recognizes our signals. The information on the television screen can also be selectively transmitted to a classical teletype terminal at our end, or ordered to come out on the line printer at the data bank centre.

The documentalist as the intermediary between the inquirer and the stored information and/or the inquirer himself can start to negotiate through the terminal with the computer processing the search on the databank. At present we have direct connection with ESRO:s (European Space Research Organisation) Computer Center in Darmstadt where about one million references are stored in following files:

Files	Number of references	From year
1. Scientific and Technical Aerospace Reports - STAR 2. International Aerospace Abstracts - IAA	510 000	1962
3. Computerized Engineering Index - COMPENDEX	105 000	1969
4. Metals Abstracts Index - Metadex	79 000	1969
5. Nuclear Science Abstracts - NSA	190 000	1969
6. Government Reports Announcements - GRA	55 000	1970
7. Electronic Components Databank	4 271	1970

Chemical Abstracts Condensates file is being tested.

The total yearly updating rating about 280 000 references.

13. CONCLUSION

During the five years of activities the documentation centre at the Royal Institute of Technology has established itself as an information centre in the fields of science and technology.
The SDI service is now well implemented and its activities are used and appreciated by scientists, research workers and engineers at the universities, research institutions and in the industrial communities.
Techniques for on-line SDI-query formulation and query alternation adaptive to user feedback are under development.
The on-line connection to the NASA:s Recon system in Darmstadt enables us to make retrospective searches in interactive mode.
Research is going on for linking up the Swedish network for Library Information system - LIBRIS - with international data banks with the objective to achieve a comprehensive information retrieval system for the whole country.

14. REFERENCES

(1) Tell, B.V., Larsson, R. & Lindh, R., Information retrieval with the ABACUS programme - an experiment in compatibility. - IAG Journal, 3(4),.1970, 323-41.

(2) Gluchowicz, Z., Selective Dissemination of Information - a transdisciplinary information retrieval system at the Royal Institute of Technology, Stockholm. - IAG Journal, 4(2), 1971, 131-48.

(3) Zennaki, M., Exposé VIRA. Paris, CNRS-CDSH Equipe Informatique, 1972. 22 p.

(4) Tell, B.V., Free text retrieval and an ordering for the printout. ENEA Tutorial Seminar on "Indexing vs. Free Text Retrieval. Studsvik 1972." 8 p. (TRITA-LIB-1037)